

National Cancer Institute



Enabling the Next Wave of Cancer Research and Care

caBIG® ANNUAL REPORT | 2008

U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health

2008 Accomplishments

- Connected 50 NCI-designated Cancer Centers into a national network, now the world's largest biomedical research "highway"
- Established the caBIG® Enterprise Support Network, with six Knowledge Centers and seven caBIG® Licensed Support Service Providers, to assist institutional and individual users in adopting or adapting their applications to join the caBIG® network
- Provided user-friendly "tool suites" for discovery and clinical research, selected from the 40+ caBIG® software applications
- Provided the information technology infrastructure for The Cancer Genome Atlas which is "powered by caBIG®"
- Collaborated with the U.S. Department of Health and Human Services and the U.S. Surgeon General's Office to develop, launch, and host the electronic version of the Family Health History tool as a foundational step towards personalized medicine
- Spotlighted caBIG® as a systems approach to the new generation of integrated research and clinical care in the U.S. Health and Human Services report on *"Personalized Health Care: Pioneers, Partnerships, Progress"*
- Connected caBIG® with the U.S. prototype Nationwide Health Information Network demonstrating national exchange of health information
- Expanded international collaborations with the National Cancer Research Initiative (NCRI) in the United Kingdom; The Beijing Cancer Hospital and the Shanghai Center for Bioinformatics Technology in China; and the King Hussein Institute for Biotechnology and Cancer in Jordan
- Applied caBIG® infrastructure as "electronic glue" to connect a new ecosystem of organizations—called the BIG Health Consortium™—to enable a 21st century model of research and care



Table of Contents

Letter from the Director, Center for Bioinformatics and Information Technology	2
---	---

Connecting the Cancer Research Community	3
--	---

Accelerating the Benefits of Personalized Medicine	12
--	----

The New World of Biomedical Collaboration	17
---	----

TO FRIENDS AND COLLEAGUES THROUGHOUT THE BIOMEDICAL COMMUNITY:

The pace of change in the U.S. biomedical enterprise has accelerated to a dizzying pace in recent months, as multiple trends – scientific, clinical, demographic, financial, technological, and societal – have converged.

We have surely entered the era of information-based medicine, in which a more precise understanding of disease origins to prevent, diagnose, treat and monitor disease in a truly individualized way for every person is now feasible, albeit not available uniformly.

The challenges, of course, remain daunting. The demographics of an aging population threaten a staggering 20 million new cases of cancer globally by the year 2030, up from 12 million per year today.¹ The tsunami of data that results from high-throughput technologies is still engulfing us. Our research institutions are still functioning for the most part as silos, not only within their own facilities but also between one another, with potentially illuminating information imprisoned inside paper records or barred by electronic systems that cannot interoperate.

Fortunately, we are making significant progress against these challenges, and informatics solutions can address many of the technical hurdles. A 21st century biomedical system is predicated on the availability of a nationwide, interoperable, interconnected information technology platform that enables information sharing. The NCI identified the need early in this decade for such an informatics initiative—unprecedented in scope for the biomedical community—and developed caBIG® (cancer Biomedical Informatics Grid) to enable rapid information exchange among previously disconnected laboratories, institutions, and sectors.

caBIG® is an enabler of individualized cancer medicine, a catalyst for change, and a model for unifying research and care in cancer and beyond. Although it is customized for cancer, it is remarkably “BIG” in potential.

As an *enabler*, caBIG® now provides the electronic infrastructure for the cancer community to conduct every step along the bench-to-bedside continuum: to manage biorepositories; to oversee clinical trials from planning through implementation through regulatory submission; to conduct molecular discovery; to integrate clinical images; and to connect researchers and institutions with a system that addresses legal, regulatory, policy, intellectual capital and contractual barriers to data exchange.

As a *catalyst*, caBIG® continually brings the need for inter-disciplinary communication and collaboration to the forefront, provoking cultural shifts and action that might otherwise be postponed. By making data easily accessible, caBIG® accelerates discovery. Beyond simply “speeding up” 20th century processes with electronic methods, caBIG® is changing the way those processes are done, and the way researchers think about their explorations.

Finally, caBIG® is serving as a *model* by which research and care can be united, so that the knowledge gained in the care setting can be harnessed to fuel research discoveries and those discoveries can be channeled back into care in a seamless virtuous cycle – called the rapid learning health care system – that uses all information as a source of knowledge for improving patient care.

The Sequencing of the Human Genome was once called a “race to the starting line.” Similarly, the unprecedented caBIG® initiative—which has engaged thousands of individuals and hundreds of institutions in the past five years—has now crossed the starting line of a new national highway of connectivity and information exchange that promises to transform biomedicine in cancer and beyond.



Ken Buetow, Ph.D.
Director, Center for Bioinformatics and Information Technology
National Cancer Institute
March 1, 2009



¹ World Cancer Report 2008, International Agency for Research on Cancer (<http://www.iarc.fr/en/Publications/PDFs-online/World-Cancer-Report>).



Connecting the Cancer Research Community

THE PRIORITIES OF THE caBIG® PROGRAM are determined by the communities it is designated to serve, which include dozens of academic and community cancer centers, thousands of scientists and clinical researchers, and countless related organizations that are “stakeholders” in the development of improved prevention, detection, diagnosis, treatment, and monitoring. The caBIG® program was designed first to connect the network of 60+ NCI-designated Cancer Centers and then to link the rest of that larger biomedical cancer community.

In 2008, 50 of the NCI-designated Cancer Centers participated in a process that assessed their biomedical informatics capabilities, their goals for biomedical informatics, and their institutional plans for implementing biomedical informatics. While Cancer Centers are highly diverse in size and scope, they have similar challenges and objectives.

Addressing the Needs of the NCI-designated Cancer Centers

The overarching goals of the Cancer Centers include achieving significant advancements in prevention, risk assessment, prediction, prognosis, biomarkers, drug development, and customized therapy for improved health outcomes, via data sharing, analysis, mining and improved regulatory and security processes. Specifically, NCI Cancer Centers seek to:

- Effectively manage the overwhelming volume of data generated by 21st century biomedicine.
- Develop informatics platforms that can leverage genomic, imaging, clinical, and population data to support basic, clinical, and population science.
- Dismantle the existing data exchange barriers to connect and streamline work done in multiple departments,

divisions, or between multiple sites in the same institution.

- Increase accuracy through automation of labor-intensive processes.
- Simplify reporting by standardizing data collection and streamlining submission mechanisms.
- Perform complex analysis across multi-dimensional, multidisciplinary data sets.
- Identify best practices in other institutions.
- Share data among NCI's Clinical Trials Research Programs.

In response to these specific needs, the caBIG[®] program has developed tools, standards, and information technology infrastructure that address many of these strategic and operational needs, including:

- Electronic management of bio-repositories from sample collection to dissemination and search/query.
- Electronic management of clinical trials

from protocol development through regulatory submission.

- Electronic management of molecular studies of genomic, proteomic, and epigenomic data and integration with clinical data.
- Electronic management of a repository of *in vivo* images and integration of images with clinical and genomic data.
- Development of standards-based data models, controlled vocabularies and common data elements for multiple scientific domains that simplify data exchange and facilitate integrative analysis across multidimensional data sets.
- Data connectivity among Cancer Centers, technically via caGrid and procedurally through the Data Sharing and Security Framework, that addresses legal, regulatory, policy, and proprietary barriers to data exchange.

caBIG[®] is Connecting NCI-Designated Cancer Centers, Community Cancer Centers and Community Oncology Programs

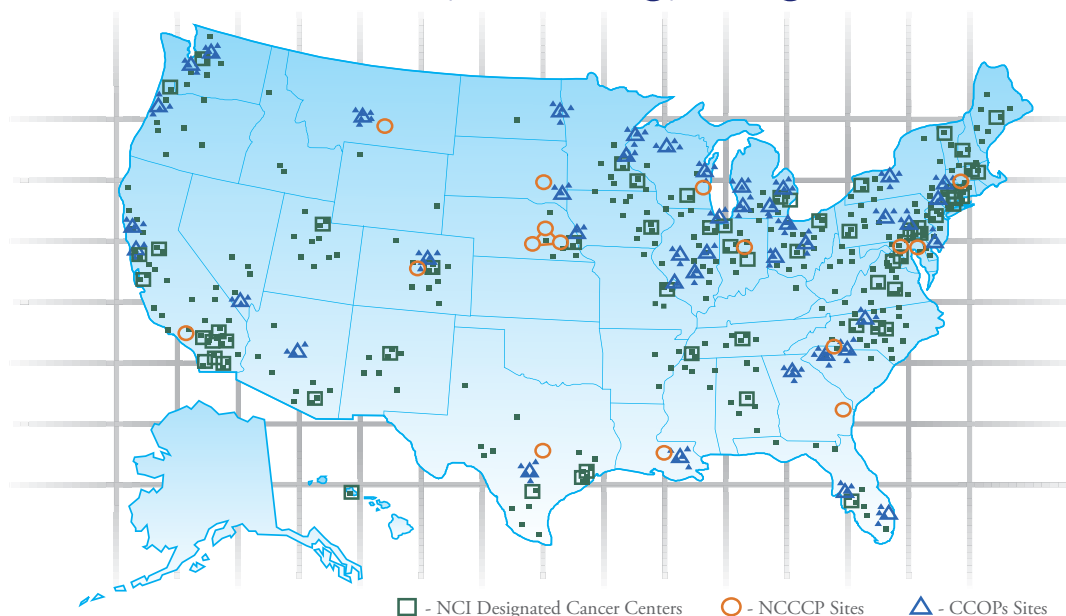


Figure 1: Cancer research and care is conducted at a large collection of institutions across the United States. A primary goal of the caBIG[®] program is to connect the 63 NCI-designated Cancer Centers and 16 National Community Cancer Center Program (NCCCP) sites, and community oncology programs into a web of cancer research and care.



Enabling Collaborative Clinical Trials

One of the ways caBIG® has empowered the NCI-designated cancer centers is by enabling collaborative clinical trials. Virtually all NCI-designated Cancer Centers use a complex collection of clinical trials management software to conduct large numbers of clinical trials—as many as 600+ different trials at a single center. caBIG® facilitates and enhances the value of these trials by providing interoperable tools that leverage standardized data collection methods to simplify data collection and analysis by:

- Eliminating duplicate data entry and reducing transcription errors
- Speeding standardized, compliant adverse event reporting via a single point of data entry
- Adopting rules-driven automated reporting to appropriate agencies internally and externally (i.e., MedWatch, IRB)
- Sharing data from clinical trials in a standard and efficient way both internally and with collaborative groups

Clinical Trials Suite. One of the ways caBIG® has empowered the NCI-designated Cancer Centers is by enabling collaborative clinical trials. This modular suite of tools (a type of Clinical Trials Management System, or CTMS) enables key aspects of clinical trials management and includes tools for

participant registration and scheduling, data storage, translation and exchange, and adverse event reporting.

More than 30 NCI-designated Cancer Centers and National Community Cancer Center Program (NCCCP) sites are currently evaluating or deploying tools from the caBIG® Clinical Trials Suite or are working with vendors of the software solutions currently in place at their center to ensure compatibility with the caBIG® suite.

Clinical Data Management Systems (CDMS).

Clinical Data Management Systems manage the collection of clinical data actually generated by a clinical trial. When compared to the paper-based systems still in use across much of the NCI community, these electronic CDMS improve the quality and comparability of data obtained at different trial sites, reduce trial administration overhead, and provide significant cost and time savings. The NCI has made strategic procurements of two CDMS available to the entire NCI-supported clinical research enterprise. The first, the Cancer Central Clinical Database (C3D), based on Oracle Clinical, provides a basic infrastructure to collect and manage clinical trial data in a manner that improves efficiency in building studies. It is currently in use at 15 Cancer Centers and at a broad collection of other trial sites.

caBIG® IN ACTION:

The **Winthrop P. Rockefeller Cancer Institute at the University of Arkansas** for Medical Sciences has developed a single sign-on clinical trials management dashboard that includes patient registration (C3PR), patient calendars (PSC), adverse events reporting (caAERS) and the LabViewer from the caBIG® Clinical Trials Suite. These tools are connected with applications for biospecimen (caTissue) management, microarray (caArray) management, and the clinical information exchange engine (caXchange) to create a comprehensive clinical management solution.

Laura Hutchins, M.D., Director of the Division of Hematology/Oncology, notes: *"We're always looking for ways to move forward as a cancer center. So caBIG® is a way to help meet some of our needs. The ultimate vision is to have a seamless integration that would allow researchers to look at clinical, tissue bank, genomic, and proteomic data in a de-identified way. We want the infrastructure to be such that the researchers can run queries on their own, ask questions, and pull information from all the different sources to get the results they are looking for."*

Her colleague, Damir Herman, Ph.D., Assistant Professor of Medicine at the Myeloma Institute for Research and Treatment, adds: *"A major goal with caBIG® is to achieve standardization and actually get a much bigger picture about cancer. There are common genetic denominators in cancer, and the sharing of data across types of cancers as a result of standardization will be advantageous. The only way to do that, I guess, would be the way we look up people in a phonebook. Currently, this [capability] is not the case [for cancer-related data], and we're hoping with caBIG®, we'll be able to achieve that."*

“By implementing caBIG®, we expect to be able to increase the research efficiency (shorter turn around time from hypothesis to results; faster discovery), translational efficiency (faster turn around time from bench to bedside; faster technology transfer), and operational efficiency of our organization.”

R. HANNES NIEDNER, M.D. | University of California San Diego
Moore's Cancer Center (MCC)

caBIG® IN ACTION:

As a consortium of several institutions across the Texas Medical Center, the **Baylor College of Medicine (BCM)**, **Dan L. Duncan Cancer Center (DLCCC)** faces the ongoing challenge of integrating its research efforts across all of its affiliates. BCM began its deployment of caBIG® with caArray, an application that makes microarray data available for researchers to see connections and correlations between genes and diseases. Since then, BCM has begun to deploy the caTissue suite and the Cancer Central Clinical Patient Registry (C3PR). David Steffen, Ph.D., Director of Informatics for BCM, notes: *“The new direction of biomedical research towards connection and collaboration makes the transition to interoperable research software inevitable. BCM has selected caBIG® because we believe that it is the enterprise that will ultimately achieve this lofty goal. We’re also finding that culture shift is as important as a shift in technology, so we are also educating our teams on the importance of sharing data with other institutions.”*

caBIG® Enables the Continuum of Molecular Medicine

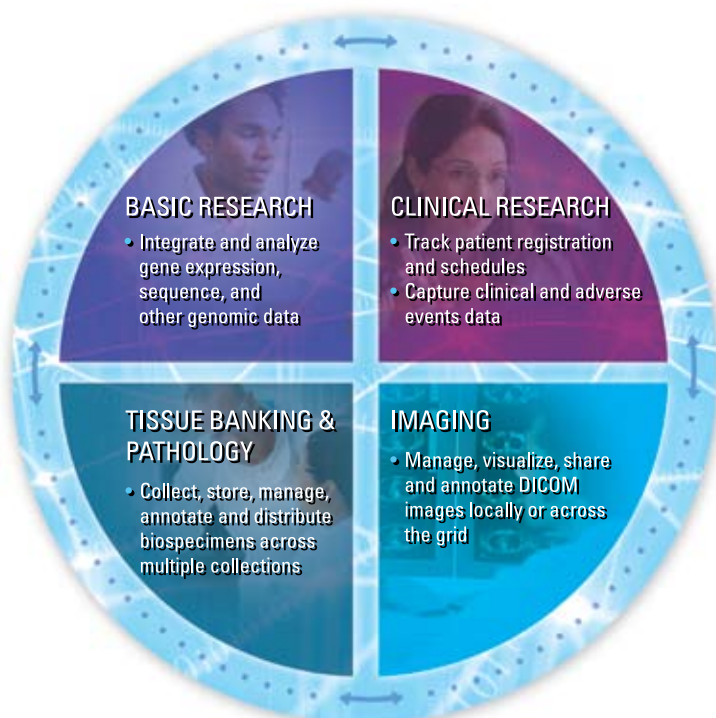


Figure 2: caBIG® enables data sharing between all critical aspects of basic and clinical research, enabling bench-to-bedside-and-back.

caBIG® IN ACTION:

Members of the NCI Community Cancer Centers Program (NCCCP) (<http://ncccp.cancer.gov>) have avidly adopted caBIG® tools and infrastructure. The NCCCP is a three-year pilot program to test the concept of a national network of community cancer centers to expand cancer research and deliver the latest, most advanced cancer care to a greater number of Americans in communities in which they live, and where 80+ percent of cancer patients are treated.

Among the members of the NCCCP, four organizations are leading the way for caBIG® implementation. The Helen F. Graham Cancer Center at Christiana Care Health System in Delaware has made significant progress toward implementing caBIG® tools, beginning with caTissue Suite 1.1. Future plans include integrating an Electronic Medical Record (EMR), and deploying the caBIG® microarray repository (caArray) and the caBIG® *in vivo* image repository (NCIA) locally before the end of 2009.

Our Lady of the Lake Regional Medical Center in Baton Rouge, Louisiana has gone live with NCIA at both its facility and at its sister site, the Mary Bird Perkins Cancer Center. Follow-up plans include using the caBIG® clinical trials data warehouse (CTODS) for patient data management and implementing the new CDMS system once available. Two centers from the St. Joseph's healthcare network are also installing multiple caBIG® applications, including NCIA and a caGrid node to support their clinical care efforts.

In response to the evolving needs of the caBIG® cancer research community, caBIG® is in the final stages of procuring the second CDMS to unify the NCI Cooperative Groups—which are responsible for over 1,700 trial sites, over 30,000 new accruals per year, and more than 100,000 patients in active follow-up—with a common, electronic standards-based infrastructure. These systems will integrate with the caBIG® Clinical Trials Suite, enabling data sharing with other caBIG®-compatible systems and tools across the NCI community.

Managing Biorepositories and Biospecimens

The caBIG® biospecimen applications (caTissue) were designed specifically to address the needs of Cancer Centers, in accordance with the NCI's Best Practices for Biospecimen Resources,² and more than 20 Cancer Centers have already installed or are planning to install caTissue to address the specimen management needs of their researchers and clinicians. The components of the caTissue Suite:

- Provide an integrated, Web-based software system that collects, annotates,

and manages specimens efficiently and in compliance with regulatory requirements

- Enable robust queries by end-users
- Identify potential tissue sources for research studies
- Simplify tissue identification for rare research tissue
- Track samples across distributed repositories
- Provide links between genomic information and clinical phenotype for specimens

Managing Genomic/Molecular Information Electronically

High-throughput genomics technologies enable novel insights into the underlying molecular etiology of cancer. In particular, microarray experiments have been instrumental in characterizing subtypes of certain cancers, leading to improved diagnosis and even guiding the course of patient treatment.

But these technologies generate overwhelming amounts of data, measured in “terabytes.” The caBIG® program addresses the challenge of managing and analyzing these large data collections by developing novel interoperable tools that help

researchers find underlying connections between diverse data sets that are difficult or impossible to find using conventional tools and methods.

Specific requirements for handling genomic data noted by the NCI-designated Cancer Centers include the ability to:

- Perform integrated analysis of gene expression, sequence, SNP, pathway, image, and protein structure
- Store, annotate, and access microarray data
- Compare data from independent studies performed at different institutions by integrating cancer data repositories

So far, about 20 Cancer Centers have adopted the caBIG® microarray software (caArray) to collect and manage their experimental gene expression data. caArray answers the researchers' needs for browser-based and programmatic access to microarray data obtained from a variety of commercial analysis platforms stored locally or that are available on caGrid. caBIG®-developed data models help integrate microarray data with other data types including imaging, sequence, and tissue information.

² NCI Office of Biorepositories and Biospecimen Research, <http://biospecimens.cancer.gov>.

caBIG® IN ACTION:

The **Wake Forest University Comprehensive Cancer Center (CCCWFU)** is committed to utilizing caBIG® tools to enhance productivity among previously isolated groups within its own organization, as well as within the larger cancer community. By deploying caAERS, a caBIG® application for adverse event reporting, CCCWFU hopes to enhance the operational efficiency of clinical trials, particularly as they engage in research collaborations with other institutions.

Bob Morrell, the Center caBIG® Deployment Lead, notes: *“By deploying caAERS, we are elevating the application from merely adverse event (AE) reporting to achieve adverse event analysis. This process will result in better standardized protocols, and it will ensure that the appropriate AE reports are submitted to regulators. The trend toward multi-institutional clinical trials necessitates a system that allows the exchange of data across multiple systems. caBIG® provides the architecture to achieve such communication, along with the tools to enable seamless coordination of trials with other institutions.”*

Supporting Medical Imaging Studies

Radiology and pathology images are critical in supporting cancer diagnosis, research, and clinical treatment. These technologies are shifting from film to digital formats, creating new data storage and management issues. Until now, Cancer Centers have lacked the technology to enable simple, secure, interoperable access to medical imaging data.

caBIG® has developed *in vivo* imaging software to support clinical trials. The Cancer Centers need imaging tools that:

- Enhance physician-to-physician communications between oncologists and radiologists
- Enhance collaboration and enable data sharing between physicians at remote locations, without costly image duplication
- Foster interdisciplinary research that integrates medical imaging techniques into basic, translational, and clinical cancer research studies

The centerpiece of the caBIG® imaging program is the **National Cancer Imaging Archive (NCIA)**, a searchable online repository of *in vivo* medical images. NCIA provides access to DICOM (Digital Imaging and Communications in Medicine) images, shared annotations, and image metadata. Working in conjunction with the Radiological Society of North America (RSNA), NCIA also utilizes the Clinical Trials Processor electronic submission tool to allow secure and de-identified electronic submission of image data.

In addition to NCIA, the caBIG® program has developed tools to support image annotation and the production of specialized analytical tools to support specific needs of researchers:

AIM: Annotation and Image Markup software provides annotation tools, associated ontologies, and object models to create, validate, and render standardized image annotations and markups.

XIP: The eXtensible Imaging Platform is an open-source toolkit that speeds development of interoperable medical imaging applications, simplifying post-processing onsite.

In total, almost 40 organizations, including seven NCI Cancer Centers, commercial organizations, and other research institutes within the federal government are using or evaluating caBIG® imaging tools to enhance their research and care efforts.





Supporting caBIG® Adoption: The Enterprise Support Network

To facilitate the use of caBIG® tools and connection to the caBIG® network, the NCI Center for Biomedical Informatics and Information Technology (CBIIT) launched the Enterprise Support Network (ESN) in 2008. The ESN has two components—**Knowledge Centers** and **Support Service Providers**—which together provide comprehensive support for organizations and individuals (including research end-users and IT staff) who want to leverage caBIG® tools, standards, and technology in their labs. Each of the six **Knowledge Centers** has a specific domain of expertise, including such diverse areas of caBIG® technology as Clinical Trials software, caGrid architecture, and Data Sharing and Intellectual Capital. **Support Service Providers** are third-party organizations licensed by the NCI to provide high quality, hands-on comprehensive technical and end-user support in areas such as software customization, documentation and end-user training, and help-desk support under fee-for-services agreements with the contracting organizations. (For ESN information, please visit <https://cabig.nci.nih.gov/esn>).

caBIG® IN ACTION:

At **Duke University's Clinical Genomics Studies Unit (CGSU)**, P. Kelly Marcom, Ph.D., is leading a Phase II Trial, Evaluating the Performance of Genomic Profiles of Chemosensitivity to Direct the Use of Preoperative Chemotherapy in Early Stage Breast Cancer, supported by the U.S. Department of Defense's Breast Cancer Research Program. This trial is one of several underway at the Duke Clinical Genomics Studies Unit testing the clinical utility of genomics guidance of therapy. The Duke team members are applying components of the caBIG® Clinical Data Management System, including C3D-based data architecture to integrate genomic signature data, electronic case report forms, and caTissue for tissue specimen tracking.



caBIG® IN ACTION:

At the **Kimmel Cancer Center at Thomas Jefferson University (TJU)**, the biospecimen collections are maintained in tissue banks dispersed across the various departments of the hospital, making it difficult for researchers to find the specific specimens they need to support their research. TJU turned to caTissue to provide a simple, user-friendly interface to manage and query their diverse biospecimen collections. Jack W. London, Ph.D., Director of the Informatics Shared Resource at the Kimmel Cancer Center, comments: *"For us at Jefferson, caTissue Suite answered a need to provide a comprehensive tool for our researchers to identify specimens required for their research. We've been able to link three internal tissue banks already, and we are expanding our capabilities this year. The ability to share this information using caGrid is very valuable to our collaborations with other institutions."*

“...caBIG[®] solves the major technical problems needed to exchange data within the cancer community, from molecules to patients’ clinical data. It is intended to revolutionize cancer research. Participants already include most of the sixty-three NCI-designated Cancer Centers in the United States.³”

LYNN ETHERIDGE | George Washington University



caBIG[®] IN ACTION:

At the **University of California at San Francisco**, Max Wintermark, M.D., Director of the NeuroCardioVascular Imaging Laboratory, and the lead investigator working with caBIG[®]-enabled imaging tools, notes: *“caBIG[®] is basically the response to what researchers like me have hoped for, for a long time. caBIG[®] can house multiple types of data, whether they are imaging or clinically oriented. Also, it is likely that in the future, other types of data will need to be stored (e.g., blood and genetic data), and caBIG[®] can accommodate these types of data. caBIG[®] enables data aggregated from multiple sites to appear as an integrated research tool set or large database while individual resources remain under the control of the originating organization. Thus, there could potentially be no significant upfront cost associated with a Center’s decision to participate in a caBIG[®]-based research system.”*

³ Etheredge, Lynn M. 2009. Medicare’s Future: Cancer Care. Health Affairs, 28(1):148-159.

caBIG® IN ACTION:

The **Jackson Laboratory Cancer Center** studies mouse genetics to develop disease models that improve the prevention, diagnosis, and early detection of cancer in humans. The Jackson Laboratory has deployed caArray to make its internal data repositories available through caGrid. Chuck Donnelly, Director of Computational Sciences, discusses how caBIG® is transforming research at Jackson Labs: *“The deployment of caArray has allowed us to create a state-of-the-art microarray analysis system for our research staff by providing a means of simplified access to data analysis tools that were previously unavailable. In addition to driving a shift in the way research is taking place at our lab, caArray has also enabled us to make our data available to other institutions. We believe this data will enable cancer research by informing pre-clinical and clinical studies in humans, essentially initiating a paradigm-shift in the way research is currently taking place.”*

Data Interoperability: Overcoming the Tower of Babel

Efficient and effective use of biomedical data is essential to understanding the molecular processes that underlie cancer and to developing better interventions. This ability to access and use data between different systems is termed “semantic interoperability.” But multiple data representations and software products have evolved over the years, seriously limiting interoperability.

caBIG® enables data interoperability by systematically adopting internationally recognized data standards where they exist and by working closely with professional organizations to develop vocabularies, define common data elements (CDEs), and extend data models where such standards do not exist or are inadequate to the task. For example, in partnership with key industry standards organizations such as the Clinical Data Interchange Standards Consortium (CDISC) and Health Level 7 (HL7), caBIG® has helped create the Biomedical Research Integrated Domain Group (BRIDG) (<http://bridgmodel.org/>). caBIG® develops tools that adhere to these standards, allowing multiple institutions to exchange data and research collaboratively, similar to the way agreed-upon interoperable standards enabled the banking system to offer the interconnected network of ATMs

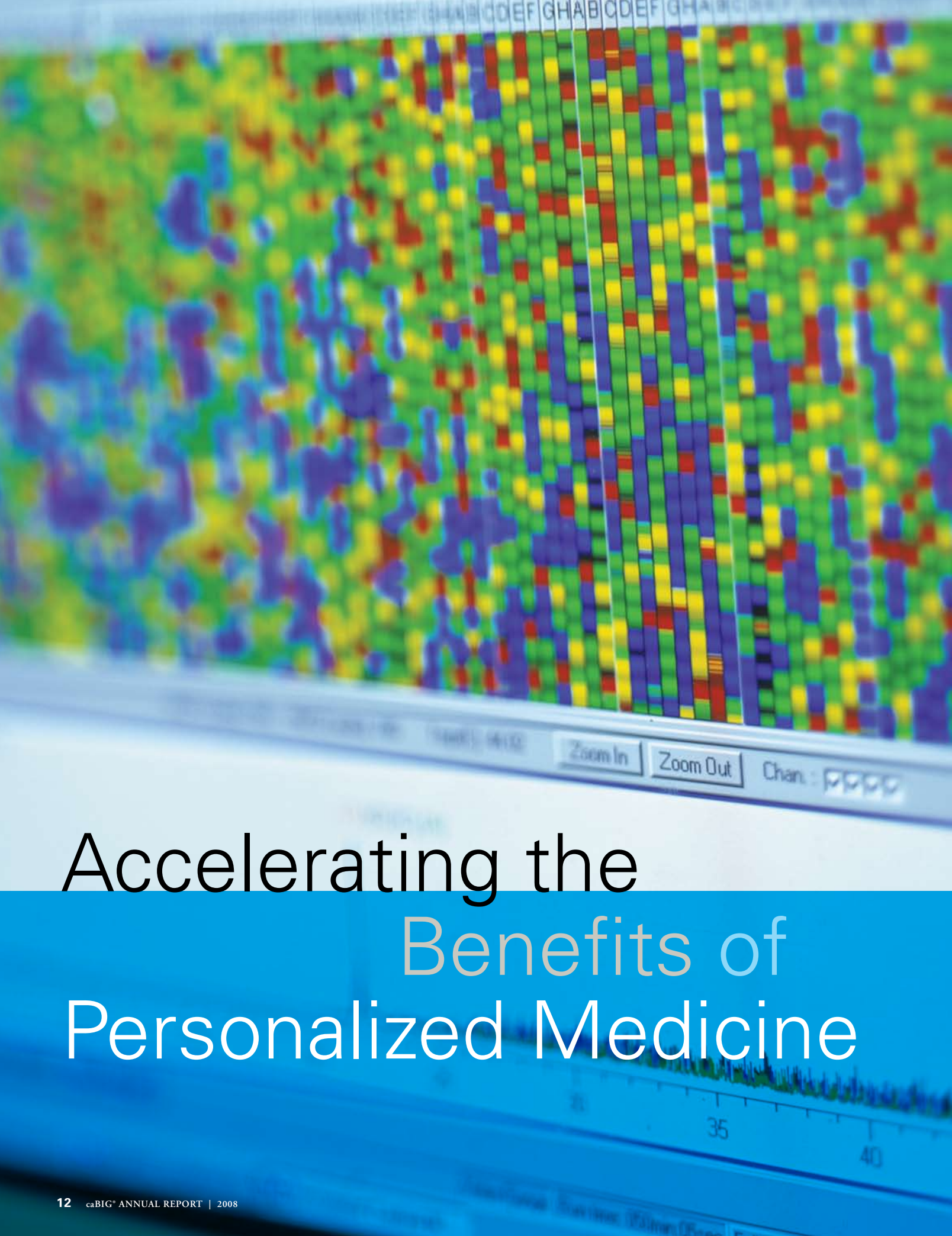
that enable us to access our money anywhere in the world, regardless of an individual bank’s location.

caBIG® provides two complementary pathways to achieving interoperability. Institutions may “**adopt**” caBIG® tools and use those applications to connect to caGrid, or they may “**adapt**” commercial or in-house developed software to be caBIG®-compatible. Many institutions pursue both approaches, depending on their specific needs and their existing IT infrastructure. A number of the NCI-designated Cancer Centers currently connecting to caGrid are working with commercial software vendors to ensure essential components of their IT infrastructure are fully-caBIG® compatible. The choice of pathway by a particular institution depends on many factors, including:

- The extent to which existing tools meet the needs of the end users or if additional capabilities are required.
- The receptiveness of end-user researchers to learn new tools and processes, or their desire to continue using familiar tools and workflows.
- The available resources for internal development or customization of existing tools to become caBIG®-compatible, versus installing existing caBIG® tools.

caBIG® has developed Application Programming Interfaces (APIs) and Software Development Kits (SDKs) that provide detailed instructions about application interfaces, simplifying the process of developing new caBIG®-compatible software, or modifying existing software to become caBIG®-compatible.

Recognizing that many organizations have a well-established IT infrastructure that satisfies some of their requirements, but still must provide interoperability within that existing infrastructure, the caBIG® program developed the caXchange and caAdapter programs to address these needs. caXchange generates messages in the industry standard HL7v3 format and simplifies the mapping and exchange of clinical trials information between applications and point-of-care systems. A Lab Viewer is also provided to assist in the inspection of clinical laboratory data. caAdapter is an open-source toolkit that integrates tightly with caXchange and provides developers with web services and APIs to simplify application integration and model mapping services between different kinds of data sources.



Accelerating the Benefits of Personalized Medicine

THE PROGRESS MADE BY THE caBIG® PROGRAM in 2008 does not stop with connecting the NCI-designated Cancer Centers and the National Community Cancer Center Program (NCCCP). A wide range of research and clinical efforts underway in the United States and internationally are “powered by caBIG®,” further accelerating fulfillment of personalized medicine.

Discovering Biomarkers

Researchers seek “biomarkers”—readily measurable characteristics, such as gene expression patterns, proteomics profiles, mutations and SNPs—that subgroup the patient’s cancer and may enable physicians to identify optimal therapies or predict the likely course of disease progression. caBIG® tools and technologies simplify the integration of multidimensional genomic and clinical data, giving researchers the power to ask and answer more complex, biologically important questions in support of large-scale scientific initiatives such as:

TARGET (Therapeutically Applicable Research to Generate Effective Treatments). The TARGET program has recently used gene expression information from microarrays and large-scale sequencing data to identify and subsequently validate several novel recurrent mutations that were

only found in acute lymphoblastic leukemia (ALL) patients with poor clinical outcomes.⁴ These mutations cause activation of the JAK gene, and the data suggest it may be possible to develop a diagnostic panel to identify patients with poor predicted outcomes. Development of a novel therapy for this group of ALL patients may also be possible. The same TARGET data used by these researchers is accessible through the caBIG®-enabled Cancer Molecular Analysis (CMA) portal (<http://cma.nci.nih.gov>) and the TARGET data portal (<http://target.cancer.gov/dataportal/>).

TCGA (The Cancer Genome Atlas Project). TCGA (<http://cancergenome.nih.gov>) is a collaborative effort between the National Cancer Institute and the National Human Genome Research Institute to evaluate systematic approaches to identifying the molecular basis of human cancer using

genome analysis technologies, including gene expression, copy number alteration, and large-scale genome sequencing. caBIG® analysis and visualization tools were used by TCGA network members to identify three novel gene mutations associated with the brain cancer glioblastoma multiforme (GBM) in 2008.⁵ The same TCGA data used by these researchers is accessible through the caBIG®-enabled Cancer Molecular Analysis (CMA) portal (<http://cma.nci.nih.gov>) and the TCGA data portal (<http://tcga.cancer.gov/dataportal/>).

CMA (The Cancer Molecular Analysis Portal). The CMA Portal (<http://cma.nci.nih.gov>), enabled by caIntegrator, exemplifies the caBIG® core principles of open development and federation by linking analysis programs developed at three different organizations in an easy-to-use Web portal. The CMA portal helps researchers correlate clinical characteristics—such as survival data

“What is required in cancer research to find definitive answers is a system to share data and leverage all the events in the cancer world. It is impossible to succeed without embracing that notion. The concept of caBIG® is therefore right on target.”

KIM LYERLY, M.D. | Director
Duke Comprehensive Cancer Center

⁴ Charles G. Mullighan et. al. 2009. Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. *N Engl J Med*, Jan 29;360(5):470-80.

⁵ The Cancer Genome Atlas Research Network. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, October 23;455:1061-1068.



and tumor staging—with genomic data from a variety of data sets to find novel correlations that would be difficult, if not impossible, to find using conventional means. All data generated through the TCGA and TARGET projects are currently available through the CMA portal and additional data sets are being added continuously.

REMBRANDT (REpository of Molecular BRAIn Neoplasia DaTa). The REMBRANDT project seeks to characterize a large number of adult and pediatric primary brain tumors and identify biomarkers by correlating molecular data with extensive retrospective and prospective clinical data. Approximately 900 cases have been examined so far, with more samples added monthly. More than 300 researchers use the REMBRANDT Web portal (<http://rembrandt.nci.nih.gov>), which provides the ability to perform *ad hoc* queries across multiple diverse data types and helps illuminate subtle differences between subclasses of brain tumors while assisting in decisions regarding patient treatment. The portal is enabled by caIntegrator and leverages the caBIG® Clinical Genomics Object Model (CGOM) to provide Web-based and programmatic access to the data.

VASARI (Visually AccesSABLE Rembrandt Images). Researchers are enhancing the

REMBRANDT data by adding clinical Magnetic Resonance (MR) images obtained on the samples from the REMBRANDT program. The VASARI project seeks to improve the classification of glioma tumors by linking MR images with histology and genetic data, thereby validating image features as effective biomarkers for the progression of the disease. The caBIG® NCIA (National Cancer Imaging Archive) provides image storage and analysis functions while caIntegrator provides the search capabilities.

CGEMS (Cancer Genetic Markers for Susceptibility). The CGEMS project represents the first public release of a GWAS study for prostate and breast cancer. Over 500,000 SNPs have been analyzed so far using the caBIG®-developed tool, caGWAS, and the results made available using caIntegrator through the CGEMS data portal (<http://cgems.cancer.gov>). CGEMS researchers have identified variations in FGFR2, a gene associated with increased risk for breast cancer, and multiple genes associated with increased risk for prostate cancer.

Empowering Integrated Clinical Trials

Adaptive clinical trials can leverage information collected during the course of the trial to direct patient treatments, potentially improving outcomes for

individual participants and reducing the overall cost and time to run the trial. caBIG® is enabling the next generation of adaptive clinical trials by providing novel, standards-based tools that address key aspects of data collection, protocol management, multisite management, and regulatory submission.

I-SPY (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular analysis). The I-SPY 1 trial is a national study to identify biomarkers that may predict response to therapy for women with late stage breast cancer. Over 300 women with stage II and III breast cancer have been enrolled to date.

caBIG® tools enable the I-SPY trial by integrating clinical, MR imaging, gene expression, CGH, immunohistochemistry, and other data types. The study has already established new standards for MR imaging and developed novel tools for data sharing, tissue tracking, common information repositories, and clinical trial automation that can benefit future trials. caIntegrator provides data warehousing and data mining access for researchers through a user-friendly Web portal.

TRANSCEND (TRANSlational Informatics System to Coordinate Emerging Biomarkers, Novel Agents, and Clinical Data). The TRANSCEND project extends the work of I-SPY 1 to further enhance the clinical trial data collection infrastructure. TRANSCEND uses Web-based case report forms (CRFs) to simplify data collection at two I-SPY trial sites and to demonstrate integration of an electronic health record system with the bioinformatics infrastructure in place for the I-SPY 1 trial. In addition to the caBIG® tools used in I-SPY 1, caTissue and NCIA are part of the informatics infrastructure being developed for TRANSCEND.

Empowering Population Science

PopSciGrid (POPulation SCIENCE Grid). The PopSciGrid was developed at Northwestern University, in collaboration with the NCI Population Sciences program, to demonstrate the ability of caBIG® grid technology to host data types completely

Enabling Data Sharing by Addressing Policy Issues

caBIG® provides both the infrastructure to enable data sharing and the tools to control data access. Issues of data security and evaluating which data sets can be shared fall to the Data Sharing and Intellectual Capital (DSIC) Workspace, which addresses the legal, regulatory, ethical, policy, academic, proprietary, and contractual issues related to data exchange for public health and research purposes. caBIG® is based on the belief that strong confidentiality, privacy, and security measures are both necessary and feasible in any electronic health information exchange environment, and that the measures can be scaled to accommodate a broad range of participants without unnecessarily impeding scientific discovery and medical progress.

caBIG® provides guidelines and tools to help researchers evaluate the sensitivity of their data. caBIG® has developed the Data Sharing and Security Framework (DSSF), which includes analytical tools to help researchers address the issues of federal privacy regulation, human participant protections, sponsor contract compliance, and proprietary interests. For information about the specific tools and support, please visit the DSIC Knowledge Center (<https://cabig-kc.nci.nih.gov/DSIC/KC>).

“There is an enormous amount of clinical data now coming out of these studies—clinical data on the patient, their outcome, and their treatment, as well as all of their demographic characteristics and then the genetic data on the tumor and the patient. It’s a large amount of data, and we need to have the tools to be able to put all of that data together to integrate it to communicate with other institutions who may be doing similar studies so that we can compare and collaborate—standardized tools that allow one center talk to another center or one investigator to talk to another and then allow the integration of this vast amount of data. And I think that’s what caBIG® is hoping to do, intending to do is to provide that standardization and flexibility to implement new techniques and new methods as they come along.”

C. KENT OSBORNE, M.D. | Director
Baylor Cancer Center

“I think that we’re entering an era where the greatest impact that we can have in terms of improving care for cancer patients is going to be through collaboration. And if we are really going to be able to collaborate, we have to be able to communicate effectively. We have to be able to share data. We have to be able to share biospecimens. It’s about the impact, not about the ownership. And I view the caBIG® initiative as a strategy really to provide a translator to support collaboration and sharing of data – sharing of information across institutions.”

MARY BECKERLE, PH.D. | Executive Director

Huntsman Cancer Institute at the University of Utah

different from traditional cancer research information. Using population science data on smoking prevalence, cigarette tax data, and geographical information at different institutions—connected with caGrid technology—researchers performed federated queries to understand the impact of higher taxes on smoking. Based on this proof-of-concept, the Population Science grid is being leveraged for a similar project to gain public health knowledge from obesity data.

Enabling the Nationwide Health Information Network

caBIG® technology is being integrated into the architecture of the Nationwide Health Information Network (NHIN) (<http://www.hhs.gov/healthit/healthnetwork/background/>), whose goal is to provide secure, nationwide access to health information by connecting researchers, caregivers, providers, and patients in a seamless network.

caBIG® software provides the “IT glue” to connect different data types and disconnected IT systems, providing physicians and researchers with access to critical healthcare data regardless of where it is stored.

Developing Electronic Health Records for the Oncology Community

Electronic health records (EHRs) are widely perceived as essential to the achievement of improved healthcare and personalized medicine. While broadly applicable EHRs are widely available, there remains a need for a specialty EHR that addresses the unique data needs and physician support requirements of oncology.

The American Society of Clinical Oncology (ASCO), the preeminent professional society for oncologists, and the NCI are collaborating to design an oncology EHR that will utilize caBIG® standards for interoperability. This new EHR will then be deployed to the participating hospitals of the National Community Cancer Center Program (NCCCP), bringing advanced information technology to traditionally underserved patients. By leveraging existing work already conducted by a variety of organizations, it is anticipated that a pilot program could be deployed to the NCCCP sites in 2009.





The New World of Biomedical Collaboration



21ST CENTURY BIOMEDICAL TECHNOLOGIES are changing more than scientific knowledge and methods—they are changing the scientific community itself and the ways in which individual scientists and scientific institutions interact cooperatively towards a common goal.

As caBIG® technology accelerates that dramatic shift in biomedical research, its impact extends far beyond the borders of the United States and even beyond cancer-specific research. caBIG® is actively collaborating to supply research-enabling technology to programs in other countries, to adapt tools and technology developed elsewhere to U.S. cancer research and clinical care, to leverage caBIG® infrastructure for research in other diseases, and to use caBIG® as an electronic “connector” to link disparate sectors of the healthcare system.

Supporting Cancer Research Internationally

The NCI and the U.K. National Cancer Research Initiative (NCRI) have a long-standing collaboration to leverage the best applications and technologies developed by each organization. The NCRI is leveraging caGrid technology for its applications while the caBIG® program is planning to adopt the NCRI Oncology Information Exchange (ONIX) portal, providing a Web-accessible entry point to the combined resources of the NCRI and the caBIG® program for the mutual benefit of cancer researchers working in the U.S. and the U.K.

In September 2008, the NCI and the NCRI held a joint conference, *Biomedical Informatics*

without Borders: Enabling Collaboration to Strengthen Research and Care, to explore ways to speed scientific discovery. At the meeting, nearly 500 technologists and researchers learned about global grid initiatives and tools that facilitate data sharing across national borders and discussed opportunities to power future collaboration.

caBIG® is also actively collaborating with the Shanghai Center for Bioinformatics Technology (SCBIT) (<http://eng.scb.it.org>) to develop mutually beneficial exchanges of knowledge and technology, with the ultimate goal of driving widespread adoption of caBIG® technology into the People's Republic of China. In addition, Duke University is partnering with the Beijing Cancer Hospital

of Peking University to conduct a Phase II clinical trial for breast cancer using the CTMS software suite.

The NCI is also supporting the development of the new King Hussein Institute for Biotechnology and Cancer near Aman, Jordan. When completed in 2011, this center will support basic and translational research and will provide state-of-the-art medical care for cancer patients in Jordan and other Middle Eastern countries. The collaboration is focused on the adoption of caBIG® to provide the standards that will enable data interoperability and on the widespread adoption of electronic health records.

caGrid: National “Pipeline” for the Cancer Community

At the heart of the caBIG® program is an IT infrastructure that enables connectivity among people, organizations, data, and analysis tools. Invisible to the end-user and customized for the specific needs of biomedical researchers, caGrid is a model-driven, service-oriented architecture that provides standards-based core services, tools, and interfaces so the community can connect to caGrid easily to share data and analysis efficiently and securely. There are currently more than 100 grid nodes online at government, academic, and commercial organizations.

For a real-time list of available grid nodes and services, visit <http://cagrid-portal.nci.nih.gov>.

Supporting Research Beyond Cancer

Designed from the outset to support the basic principles of biomedical research and clinical care, most caBIG® tools are widely applicable beyond cancer. Examples include:

CVRG (CardioVascular Research Grid). The goal of the CVRG project (<http://www.cvrgrid.org/>) is to enable collaboration and shared discovery across the cardiovascular research community. The CVRG is currently under development at Johns Hopkins University, The Ohio State University, and the University of California at San Diego. Still in the early stages of a four-year grant from the National Heart, Lung and Blood Institute, the development of standardized vocabularies for describing biomedical data, models, and data analysis applications in cardiovascular research is already in progress. Since the CVRG grid is based on the same underlying technology used in caGrid, the two independent research networks could be connected if desired, building a “Grid of Grids.”



“The Cardiovascular Research Grid will enable us to assemble large, geographically scattered research teams and bring together the leading experts in the world to focus on a common problem, regardless of their location. This grid will enable experimentalists to share their data with computational scientists who will analyze and model the data. The computational scientists will then share their results with their experimental colleagues, who will use it to refine their experiments. In this fashion, we believe the creation of the Cardiovascular Research Grid will accelerate the discovery of new approaches for treating heart disease.”

RAIMOND WINSLOW, PH.D. | Director
Institute for Computational Medicine at John Hopkins

The BIG Health Consortium: New Ecosystem for Biomedicine

An increasing number of academic medical centers and healthcare providers are reconfiguring their organizations and systems to deliver truly personalized care. But no systematic, national endeavor yet exists to connect all of the requisite constituencies and capabilities together into a seamless, network process to demonstrate the feasibility and value of this new model for healthcare.

To address this challenge, the NCI has launched the BIG Health Consortium™ as a partnership comprised of all the key stakeholders in healthcare who agree to work together in a new ecosystem.

BIG Health participants include NCI-designated Cancer Centers, NCI Community Cancer Centers, integrated healthcare providers, academic centers, medical schools, diagnostic laboratories and

product developers, personal genomics firms, large and emerging IT companies, patient advocacy and action-tank organizations, venture capitalists, biopharmaceutical companies, and government programs.

The goals of BIG Health are to:

- Demonstrate feasibility of implementing a new model of biomedicine
- Create an ecosystem of participants that seamlessly integrate research, care delivery, and consumer health information
- Break down traditional silos that are barriers to rapid discovery and learning
- Accelerate and enhance research productivity and improve clinical outcomes

caBIG® has already demonstrated that it can link the world of cancer research through the “electronic glue” of its interoperable tools and technology. caBIG® provides infrastructure for creating, communicating, and sharing bioinformatics tools, data, and research

results while using shared applications, shared data standards, and shared data models, all operating on caGrid. The infrastructure of caBIG® is being expanded and applied to connecting basic and translational science with healthcare—thus becoming “BIG” (Biomedical Informatics Grid) to underpin the working of this new Consortium.

BIG Health is focused on two major areas in which the new ecosystem can produce major benefits in eliminating the *deficiencies and inefficiencies* of our current biomedical enterprise and improving patient outcomes. The first area of focus is to demonstrate a “learning healthcare system” in which aggregated clinical outcomes are used to detect patterns of disease and therapeutic response. The second area is “virtual clinical research” to apply to molecular subgroupings to advance development of new therapies without re-inventing clinical trial infrastructure.

...

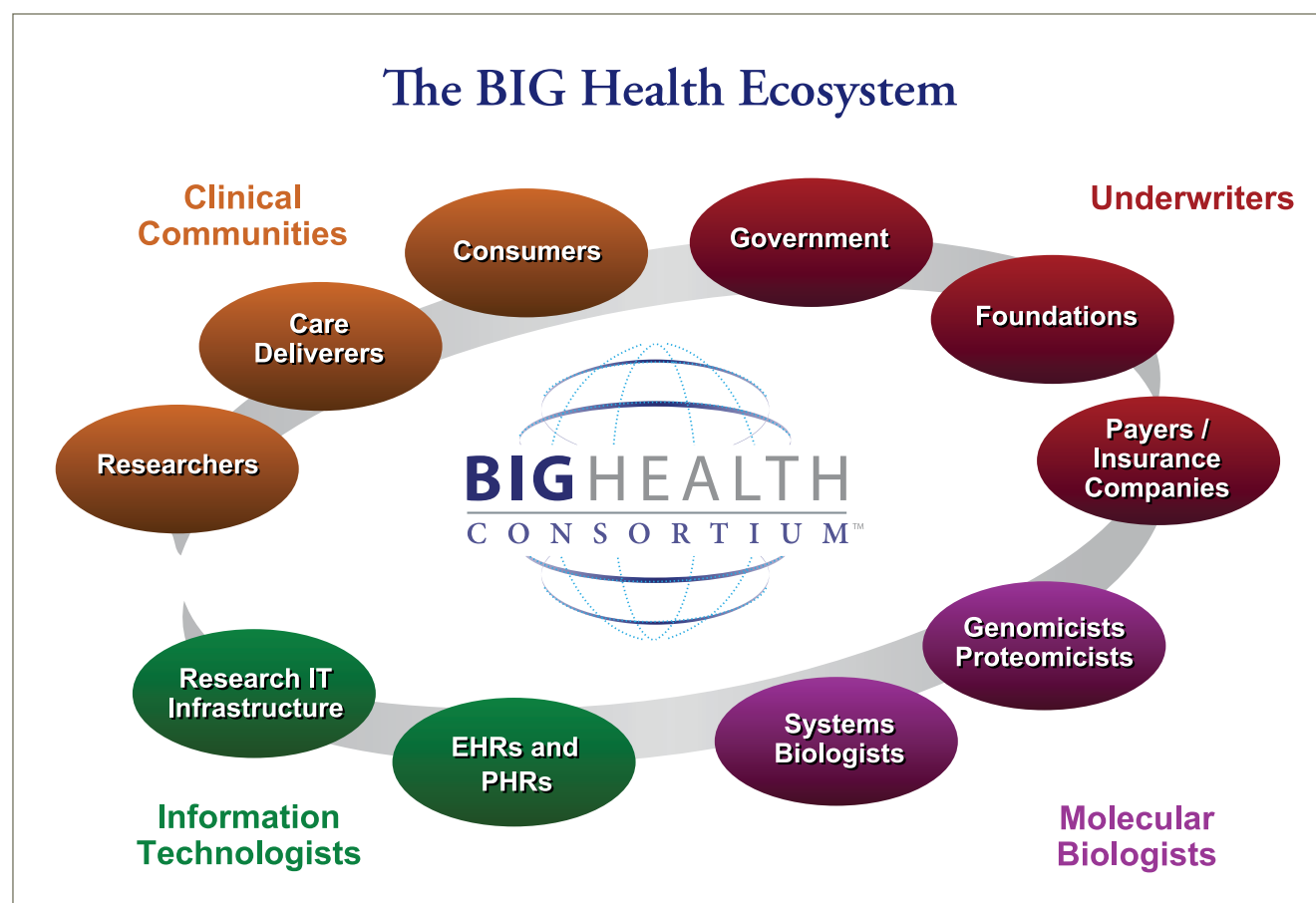


Figure 3: The BIG Health ecosystem encompasses all stakeholders in the biomedical community to foster new collaborations and models for discovery, development and delivery of care.

At the first caBIG® Annual Meeting in 2004, 170 attendees came to see and learn about a new and largely unknown endeavor. At the Annual Meeting in 2008, there were well over 1200 attendees, reflecting the growing awareness of caBIG® throughout the cancer research community and around the globe. In the future, it may not be possible to count the participants, the developers, or even the adopters of caBIG® so precisely, since widespread use will also mean that caBIG® tools are adapted and embedded in other infrastructure that is invisible to users. The connectivity and data sharing that caBIG® makes possible, however, is what truly counts. The value of caBIG® will be measured ultimately by its role in the transformation of biomedicine.

For general information about the caBIG® program,
please visit: <http://cabig.cancer.gov/>.

For technical information about caBIG® tools and infrastructure,
please visit: <https://cabig.nci.nih.gov/>.



NATIONAL[®]
CANCER
INSTITUTE

NIH Publication No.
Printed April 2009



P 1 9 0